

# **COSMO*sim3D* and COSMO*sar3D***

## **User's Manual**

COSMOlogic GmbH & Co. KG  
Imbacher Weg 46, 51379 Leverkusen, Germany

Introduction .....	3
Prerequisites .....	4
Technical details.....	4
Installation .....	7
Running COSMOsim3D .....	7
With input file .....	7
Pair-wise similarity .....	8
Keywords .....	9
General .....	9
Program Control.....	9
Random loop control.....	9
Print / Write Options .....	10
Self-consistent alignment mode .....	11
COSMOsar3D .....	11
Output.....	11
Example.....	13
COSMOview visualization .....	14

## Introduction

The Conductor-like Screening Method for Realist Solvation (COSMO-RS) has become an efficient and versatile tool for the prediction of a broad range of fluid phase thermodynamic properties based on quantum chemical calculations for solutes and solvents. In the framework of COSMO-RS, the COSMO surface polarization charge density  $\sigma$ , and its molecular surface histograms, the  $\sigma$ -profiles, have been proven to be excellent descriptors for the quantification of the most important kinds of molecular interactions in the liquid phase, such as polar interactions, hydrogen bonding and hydrophobicity. The superior suitability of  $\sigma$  for the quantification of hydrogen bond interactions has been further confirmed in a recent quantum chemical study<sup>1</sup>.

Since the same intermolecular interaction modes which govern fluid phase thermodynamics also determine ligand-receptor-interactions, it is most plausible that a  $\sigma$ -profile based description of ligand-ligand similarity or ligand-receptor-interactions should be very promising. Based on these considerations the COSMOsim<sup>2</sup> method was developed which measures ligand-ligand similarities using the molecule-specific global  $\sigma$ -profile disregarding the spatial distribution of the polarization charge density. This approach was shown to provide useful discrimination of bioisosteric and random ligand pairs, especially for smaller molecules. Besides speed, one of the major advantages of COSMOsim is that it naturally supports scaffold hopping by using the molecular COSMO-RS sigma surface instead of the molecular structure. Furthermore, analogy-based QSPR based on COSMOsim delivers powerful models for properties that are mainly governed by isotropic interactions, like logS, logP, logBB, etc.

However, the selective binding of ligands to receptors is known to be based on multiple strong interactions, and the 3D arrangement of the interaction sites of a ligand thus plays a crucial role. As a result, in target-dependent problems, one-dimensional COSMOsim tends to retrieve false-positives along with true-positives. This results from the fact that the global  $\sigma$ -profiles do not contain any information about the spatial distribution of the polarization charge densities on the molecular surface. In order to overcome this deficiency, the molecular  $\sigma$ -surface is projected onto a 3D grid in COSMOsim3D, yielding local  $\sigma$ -profiles on the grid points. Arranging molecules with maximum overlap of the local  $\sigma$ -profiles on the grid leads to optimal alignment and COSMOsim3D similarity. Based on the experience with one-dimensional COSMOsim, COSMOsim3D can be expected to enable scaffold hopping and structure-independent alignment in a natural way, since the alignment is based on the local molecular  $\sigma$ -surface instead of the molecular structure.

The array of local sigma profiles (LSPs) can not only be used for alignment and as a similarity measure as in COSMOsim3D but also as a novel set of molecular interaction field (MIF) descriptors which is optimally suited for comparative molecular field analysis, e.g. to generate a ligand-based model for the prediction of pK<sub>i</sub>. COSMOsar3D is an extension of the

COSMOsim3D method which uses the local sigma profiles of a set of aligned ligands as descriptors for 3D-QSAR analysis.

Note that in this documentation COSMOsim3D and COSMOsar3D are documented together since both methods are implemented in a single software. However, some options require a license extension for COSMOsar3D and are therefore restricted.

## Prerequisites

COSMOsim3D requires COSMO- $\sigma$ -surfaces of target and probe molecules. COSMO files can be obtained from quantum chemical calculations using quantum chemical programs with suitable COSMO implementations. Alternatively, CF-COSMO files (with the extension `.fcos`) generated by COSMOquick or COSMOfrag can be used. CF-COSMO files provide approximate COSMO  $\sigma$ -surfaces based on the 3D-geometry of the molecule and the COSMOfrag database of precalculated  $\sigma$ -profiles. For ionic compounds, CF-COSMO files for different conformers can be generated from a COSMO file of one conformer of the molecule. COSMO and CF-COSMO files contain information about the positions ( $x_i, y_i, z_i$ ), the areas ( $a_i$ ), and the COSMO polarization charges ( $q_i$ ) for all COSMO surface segments.

## Technical details

In one-dimensional COSMO-RS methods like COSMOtherm or COSMOsim, a locally averaged polarization charge density  $\sigma_i$  is calculated for each segment  $i$  according to the standard procedure<sup>3</sup>, using an averaging radius of 0.5 Å. Standard  $\sigma$ -profiles are generated as histograms with  $\sigma$ -bin width of  $\delta^\sigma = 0.1$  e/nm<sup>2</sup> from all segments of a molecule. For the generation of a  $\sigma$ -profile the area of each segment  $i$  is associated to the two neighboring  $\sigma$ -grid centers of the actual value  $\sigma_i$ , denoted as  $\sigma_{i+}$  and  $\sigma_{i-}$ , according to  $\sigma$ -distance weights

$$\begin{aligned} w_{i+}^\sigma &= (\sigma_{i+} - \sigma_i) / \delta^\sigma \\ w_{i-}^\sigma &= (\sigma_i - \sigma_{i-}) / \delta^\sigma \end{aligned} \quad (1)$$

This weighting generates a smooth and charge conserving assignment of the COSMO surface segments to histograms, ensuring that the integral of the  $\sigma$ -profile is the total surface of the molecule, and that the  $\sigma$ -weighted integral, i.e. the first moment of the histogram, is the sum of the original COSMO charges, i.e. the negative of the total charge of the molecule.

Instead of generating just one such one-dimensional  $\sigma$ -profile for the entire molecule, in COSMOsim3D a local one dimensional  $\sigma$ -profile is generated at each grid point of a regular 3D grid, i.e. a 4-dimensional histogram, with three Cartesian dimensions ( $x, y, z$ ) and  $\sigma$  as fourth dimension. For the 16 neighboring grid points of a segment  $i$  with coordinates ( $x_i, y_i, z_i, \sigma_i$ ) in the 4D space, the weights  $w^x, w^y, w^z, w^\sigma$  are computed in complete analogy to eq. 1, and the segment area is assigned to the 16 neighboring grid points according to the product of the four weights. This weighting ensures a smooth linear cross-over of the segment assignment if a surface segment is moved between the grid points.

In the first step of the COSMOsim3D procedure, the target molecule is moved to the center of the cubic grid, followed by the calculation of the local sigma profiles (*sp3d1*) on the grid points. Next, the probe molecule in question is centered and the local sigma profiles (*sp3d2*) on the grid points are calculated. Then the 3 dimensional  $\sigma$ -similarity SMS3D is calculated as a weighted sum over the  $\sigma$ -similarities of the local  $\sigma$ -profiles according to

$$SMS3D(sp3d1, sp3d2) = \frac{\sum_{ix, iy, iz} (a1(ix, iy, iz) + a2(ix, iy, iz)) sms(sp3d1(ix, iy, iz), sp3d2(ix, iy, iz))}{\sum_{ix, iy, iz} (a1(ix, iy, iz) + a2(ix, iy, iz))} \quad (2)$$

where  $a1(ix, iy, iz)$  and  $a2(ix, iy, iz)$  are the total intensities of the local target and probe  $\sigma$ -profiles at the grid point, respectively. The denominator is identical to the sum of the total surface areas of target and probe. The SMS (sigma-match similarity) calculations are performed according to reference 2, with the default parameters derived in that paper.<sup>4</sup> A target focussed or probe focussed similarity can be calculated by using the area of the target  $\sigma$ -profile or probe  $\sigma$ -profile, respectively, for normalization.

Starting from a number of determined and random orientations, the position of the probe molecule is optimized until maximum similarity is reached. This is done by a trial-and-error line search in the direction of each of the 3 unit-translation and unit-rotations, with minimum steps of 0.01 Å and 0.1°, respectively. After each translational or rotational step the SMS3D is re-evaluated and the step is accepted if it leads to an increase of similarity. The first 21 starting points in this procedure are determined, the following are random orientations. In order to achieve maximum performance of COSMOsim3D routine, for each start position of the probe the optimization of the probe vs. the target is first performed on a grid with 3-fold grid distance, which reduces the number of grid points for the SMS evaluation by a factor 27. After rough convergence on this crude grid, the optimization is continued on the fine grid if the achieved similarity is higher than a certain percentage (70%) of the maximum similarity achieved on the crude grid before, and if the optimized position is sufficiently distant from previous final points of crude grid optimizations.

Beyond the basic pair-wise alignment and similarity calculations, COSMOsim3D can also be used to build a self-consistent model from several diverse molecules which can then be used as a model target for the alignment of the probe molecules. For such a self-consistent model the first  $n$  molecules of a list of  $m$  molecules are selected as target compounds. First, a standard COSMOsim3D alignment for the first two molecules is carried out. Then the third molecule is aligned in a way to achieve maximum similarity with the averaged local sigma profiles of the previous compounds. This procedure is continued to the  $n$ -th molecule. After this first cycle is finished, the program starts again with the first compound and re-aligns it in order to achieve maximum similarity with the following  $n-1$  molecules. The same is done for the other  $n-1$  molecules. The self-consistency procedure is stopped when the total average

similarity in a loop has converged up to a predefined convergence criterion. Finally the remaining  $m-n$  compounds (probes) are aligned in order to achieve maximum average similarity with respect to the averaged local sigma profiles of the first  $n$  molecules (the "model target"). This procedure results in a consistent alignment of the entire set of compounds.

Since the self-consistent alignment can be biased by the alignment of the first two molecules, a super-self-consistent mode can be selected, in which the algorithm uses each of the first  $n$  molecules once as a start compound for the self-consistent procedure. The optimal alignment, i.e. the one with the maximum average alignment on the  $n$  molecules, is then chosen as a model to which the remainder of the list, i.e. the probe molecules, are aligned.

## Installation

The COSMOsim3D and COSMOsar3D distribution contains the executable program and some examples in a zip archive. Extracting the zip archive `cosmosim3d.zip` will automatically create the following file directory tree:

/installation directory/COSMOsim3D	binLinux32/cs3d	Linux executables
	binLinux64/cs3d	
	binWindows32/cs3d.exe	Window executables
	binWindows64/cs3d.exe	
	docs/	Documentation
	examples/	Example input and output files for various options

To install COSMOsim3D and COSMOsar3D on a Linux or Windows system, copy the appropriate executable to a local directory and set the path.

If using Bourne Again Shell (bash) on a Linux 64-bit system, add to your `.bashrc`-file:

```
export PATH=$PATH:/software/COSMOsim3D/binLinux64.
```

In a Windows environment, set the path from the system control. From the Windows "Start" menu, select "Control Panel", then "System". Select the "Advanced" tab. With the "Environment Variables" button at the bottom of the window you get a dialogue where you can edit several variables. Edit the "Path" variable from the lower part and add the path to the directory where the COSMOsim3D executable `cs3d.exe` is located. If the path is set correctly, you should be able to call COSMOsim3D on the DOS shell from any directory by typing `cs3d.exe`.

If the path for the COSMOsim3D executable is not set, you can still run a COSMOsim3D calculation by typing in the absolute path for the executable, e.g.

```
C:\Program Files\COSMOlogic\COSMOsim3D\binWin32\cs3d.exe.
```

## Running COSMOsim3D

Running `cs3d` without options will print a help message with a list of keywords.

### With input file

Execution of COSMOsim3D with a text input file allows for alignment of a list of molecules with the first molecule in the list and also for a self-consistent alignment using up to 9 target molecules. A formatted input file consisting of two parts is required:

- Three global command lines.
- List of `.cosmo` / `.fcos` files.

The number of target molecules is expected as the first entry in the first line with keyword `NSC=` and is mandatory. With `NSC=1` a single molecule is used as template for the alignment. For a self-consistent alignment, *COSMOsim3D* can use up to 9 target cosmo files (`NSC=9`) to generate an averaged target sigma profile. A super-self-consistent alignment is done when the number of molecules used to start a self-consistent alignment is indicated with keyword `NSCSTART=`. The path for the cosmo files is indicated with keyword `fdir=`. Additionally, the command lines may contain any of the keywords listed below. Lines starting with `#` are ignored.

Example input file for a super-self consistent alignment using 3 target molecules, where each of the target molecules is used as start molecule for a self-consistent alignment cycle. A local sigma profile file without compound names is written (only with *COSMOsar3D* license):

```
NSC=3 NSCSTART=3
NRAND=100 WRTLSP NOLSANAM
fdir=compounds
mol_18.cosmo
THIOL_14.cosmo
mol_09.cosmo
COO_23A.cosmo
SQ29852_2A.cosmo
SQ29852_2T.cosmo
THIOL_20A.cosmo
...
...
```

With input file `file.inp`, *COSMOsim3D* is started by

Linux systems: `cs3d file.inp`

Windows systems: `cs3d.exe file.inp`

## Pair-wise similarity

The similarity of pairs of cosmo files can be determined on the command line. The general syntax is:

Linux systems: `cs3d file1 file2 [KEYWORD]`

Windows systems: `cs3d.exe file1 file2 [KEYWORD]`

`file1` = target compound (`.cosmo` or `.fcos` file)

`file2` = probe compound (`.cosmo` or `.fcos` file)

`KEYWORD` is optional. If no keywords are given, the default values will be used. For a list of keywords, refer to section “Keywords”.

Example for the execution of *COSMOsim3D* with cosmo files for two compounds on the command line.

```

> cs3d benzene.cosmo toluene.cosmo
STATUS (cmpd,sc-cycle,ssc-cycle): 1 1 1
STATUS (cmpd,sc-cycle,ssc-cycle): 2 1 1
final similarities in_grid 0.737 0.817 0.707 0.833 0.942 0.833 9 9
benzene_toluene.inp toluene.cosmo
COSMOsim3D stops after 2 compounds

```

## Keywords

List of keywords with default values in [ ], some applicable with COSMO*sar3D* license only:

### General

**LICENSEDIR=xxx** sets path for the directory containing the license file `license.ctd`.  
Note that the path has to be set in ' or ". If omitted, the license file is searched for in the \$PWD and \$HOME directories.

**FDIR=xxx** sets path for the directory containing the `.cosmo` / `.fcos` files of the compound list.

### Program Control

**TARGETFOCUS** sets target focussed SMS3D for optimization.

**PROBEFOCUS** sets probe focussed SMS3D for optimization.

**SMSEXP=xxx** raises the SMS values calculated on the grid points to the power of xxx.

**MIN1D=xxx** sets minimum 1D-similarity to continue. [0.10]

**GRIDS=xxx** sets grid spacing in Angstrom. [1.00]

**GRIDLSP=xxx** sets LSP / LSM grid spacing in Angstrom. [1.00]

**IDELSIG=i** scaling factor for sigma intervals in LSP files. [6]

**DROT=xxx** sets angle resolution in degree. [0.10]

**DTRANS=xxx** sets translational resolution. [0.01]

**ALLOW\_INV** enables inversion steps (may destroy stereo chemistry).

**SETORI2FIRST** moves the origin to the center of the first compound. Applies only when fields are written with WRTLSM/WRTLSMA/WRTLSP/WRTLSPA keywords.

**TDIR** triggers the creation of a `cs3d_YYYYMMDD_HHMMSS_XXXXXX` folder in the current working directory; all output files will be stored there.

**TDIR=xxx** allows for selecting a specific xxx folder for output files, which will be created if it does not exist.

### Random loop control

**NSC** number of target compounds used in the self-consistency cycle [1].  
**NSC>1** triggers the self-consistent alignment mode.

**MINPS=xxx** sets minimum relative similarity of current optimum to continue with fine grid search for random loop. [0.50]

MINGD=xxx	sets minimum geometric distance relative to previously considered optima to continue random loop. [0.30]
NRAND=nnn	integer specifying number of random trials. [9]
NRAND=0	starts from initial geometry only, no random loop.
NRAND=-1	leaves the probe molecules unchanged.
SETTR=dx,dy,dz,dphix,dphiy,dphiz with NRAND=-1	moves probe by dx, dy, dz vectors and rotates probe along the axis given by dphix, dphiy, dphiz. The norm of the vector dphix, dphiy, dphiz determines the rotation angle (in degree).
SETBOX=sx,sy,sz,nx,ny,nz with NRAND=-1	sets the box start coordinates and the number of grid points for each direction.
SETORI=orix,oriy,oriz with NRAND=-1	keeps the molecules and sets the origin.
NRAND=-2	centers first (target) compound and moves all others by the same shift.
NRAND=-3	moves the origin to the center of the first compound and leaves the compounds unchanged.
NRAND0=nnn	integer specifying start of randomized trials, neglects the initial geometry. [0]

## Print / Write Options

PRINTRAN	triggers printing of random loop results.
STORERAN	triggers writing of random loop probe geometry.
WRTCOS	triggers writing of a translated/rotated probe COSMO file.
CENTER	applies only for WRTCOS keyword
WRTSDF=xxx	triggers writing of one SDF file for all compounds; new coordinates are replaced in the template xxx. By default, COSMO files and molecules in the SDF or MOL2 template are matched by coordinates. Structures in the template sd file should be the same as those you want to write to the new sdf. Also, the order of the atoms in the sd / mol2 file should be identical to the order of atoms in the .cosmo file. Hydrogen atoms are required to be present in the template sd file. If no match is found for the atoms then COSMOsim3D tries to match by molecule names.
WRTMOL2=xxx	triggers writing of one MOL2 file for all compounds; otherwise identical to WRTSDF.
REORDERCF	triggers writing of molecules in the SDF or MOL2 file in the same order as in the xxx template, rather than in the order of input COSMO files. Requires WRTSDF or WRTMOL2.
WRTLSP	writes a local sigma profile file named <compoundname>.lsp where compoundname is the name of the .cosmo / .ccf / .fcos file as used in the input file or on the command line.

WRTLSPA	writes all local sigma profiles to one file named <code>&lt;inputfilename&gt;.lsp</code> where <code>inputfilename</code> is the name of the input file.
WRTLSM	writes a local sigma moment file named <code>&lt;compoundname&gt;.lsm</code> where <code>compoundname</code> is the name of the <code>.cosmo</code> / <code>.ccf</code> / <code>.fcos</code> file as used in the input file or on the command line.
WRTLSMA	writes all local sigma moments to one file named <code>&lt;inputfilename&gt;.lsp</code> where <code>inputfilename</code> is the name of the input file.
NOLSANAM	switches off writing of compound names in <code>.lsp</code> and <code>.lsm</code> files. Applies only with one of WRTLSP, WRTLSPA, WRTLSM, or WRTLSMA.
WRTSMS	triggers writing of a local sms file. The <code>.sms</code> file is required for visualization of molecular similarity with COSMOview.
COMPRESS={ZIP GZ}	triggers writing of <code>.lsp</code> , <code>.lsm</code> and <code>.sms</code> files in ZIP or GZ format.
OUTFIL	triggers writing of the translated/rotated geometry as <code>CS3D.xyz</code> .
OUTFIL=xxx	sets the filename for the results. By default the format is xyz. The extension pdb is optional.

### Self-consistent alignment mode

NSC	number of target compounds used in the self-consistency cycle [1]. NSC>1 triggers the self-consistent alignment mode.
NSCSTART	number of compounds used to start self-consistency cycle [1]. NSCSTART>1 triggers the super-self-consistent alignment mode.
DELSCCONV	convergence criterion for average similarity in self-consistency mode [0.01].
KSCMAX	Maximum number of self-consistency loops [5].

### COSMOsar3D

Molecular fields for COSMOsar3D are written with keywords WRTLSP (WRTLSPA) and WRTLSM (WRTLSMA) and can be used for subsequent PLS analysis. The NOLSANAM keyword switches off writing of compound names in `.lsp` and `.lsm` files, so that the files can be used directly for PLS in the *open3dqsar* program<sup>5</sup>.

### Output

In general, all output is written to the working directory except when writing to a subdirectory is triggered with the TDIR keyword.

If COSMOsim3D is run with input file `file.inp`, an output file `file.out` will be written in addition to the standard output which is printed to the screen. The output file starts with information about the program version and license. The keywords from the input file are printed to a separate line. In the case of a super self-consistent alignment, the name of the start

compound yielding the best averaged similarity of the target compounds is also printed to a separate line. In the following, one line per molecule is printed to the output file.

- number of the current compound in the compound list of the input file.
- six different similarity values per compound:
  - SMS3D 3D sigma similarity as calculated by COSMOsim3D, eq. (1)
  - TARGETSMS3D target focussed 3D similarity
  - PROBESMS3D probe focussed 3D similarity
  - SMS1D global molecular similarity
  - TARGETSMS1D global target focussed molecular similarity
  - PROBESMS1D global probe focussed molecular similarity
- integer value indicating the number of the random start orientation leading to the optimal alignment.
- integer value indicating the lowest number of a random start orientation leading to an alignment with an SMS value within a certain threshold of the best SMS3D value found.
- sc-comp for target molecules, in\_grid/mol\_out for probe molecules (indicates if the molecule is inside or outside the 3D grid).
- file name of the current molecule.

The standard output, usually printed to the screen, contains a status line and a similarity line per compound.

Entries in the status lines are the number of the current compound and the numbers of the self-consistency cycle and super-selfconsistency cycle.

Entries in the similarity lines are:

- in\_grid/mol\_out indicates if the molecule is inside or outside the box.
- six different similarity values per compound, as in the output file.
- integer value indicating the number of the random start orientation leading to the optimal alignment.
- integer value indicating the lowest number of a random start orientation leading to an alignment with an SMS value within a certain threshold of the best SMS3D value found.
- Input file name.
- File name of the current probe molecule.
- Keywords from the input file.

In the case of a self-consistent or super self-consistent alignment the standard output written to the screen contains a protocol of the complete alignment procedure. There are several blocks corresponding to the cycles in the self-consistent run. The self-consistency procedure is finished when a line stating "self-consistency achieved !" is printed.

After the self-consistency is achieved, the compounds which are not part of the self-consistency cycle, i.e. the probe molecules, are aligned and the similarities to the averaged sigma-profiles from the self-consistent procedure are printed.

Additional output files are:

#### `_tr.cosmo` files

The `_tr.cosmo` files contain the aligned (translated and rotated) molecules, i.e. structures and sigma surfaces. Writing of the `_tr.cosmo` files can be triggered explicitly with the `WRTCOS` keyword. Since some other options require the `_tr.cosmo` files for further processing, the `_tr.cosmo` are also written without the `WRTCOS` keyword if another option (e.g. `WRTLSPA`) requires this.

#### `.sms` files and `_tr.sms` files

Writing of the `.sms` files is triggered by the `WRTSMS` keyword. The `.sms` files contain the local similarity information in a way which can be visualized with *COSMOview*: The local values of the *COSMOsim3D* similarity are used to control the COSMO surface transparency. By default, regions with high similarity are shown with low transparency and thus have intensive colors, while areas with low similarity get high transparency, so that mainly the ball & stick structure of the molecule appears. Note that visualization of molecular similarity with *COSMOview* requires `_tr.cosmo` and either `.sms` or `_tr.sms` files.

#### `_all.sms` files

In the case of a self-consistent alignment using more than one target compound, `WRTSMS` will also write a file named `file_all.sms`, where `file.inp` is the name of the input file. In this file, the averaged similarity of the target molecules is used to control transparency of all molecules. This may be useful to identify relevant regions in molecules for binding.

#### `.lsp` files and `_tr.lsp` files

`.lsp` files contain the local sigma-profiles computed by *COSMOsim3D*. Writing of `.lsp` / `_tr.lsp` files per compound (`compoundname.lsp`) is triggered by the `WRTLSP` keyword, while `.lsp` files for the complete compound list (named `file.lsp` using the name of the input file `file.inp`) are written with the `WRTLSPA` keyword. Note that when `WRTLSP` or `WRTLSPA` keywords are used, `_tr.cosmo` files will be written additionally.

#### `.lsm` files

Similar to the `.lsp` files, *COSMOsim3D* can write files containing the local sigma-moments. `.lsm` files can be written per compound (`compoundname.lsm`) with the `WRTLSM` keyword, or, alternatively, for the complete compound list (`file.lsm`) with the `WRTLSPA` keyword. Both keywords also trigger writing of `_tr.cosmo` files.

## Example

In the following, an example procedure for the alignment of a data set of THR ligands is described in detail. In the subdirectory `examples` of the *COSMOsim3D* installation directory, a prepared input file named `thr-align.inp` can be found.

```

NSC=1
NRAND=50 WRTSMS
fdir=thr-ligands
45.cosmo
1.cosmo
2.cosmo
3.cosmo
4.cosmo
5.cosmo
6.cosmo
...

```

With this input file, an alignment using one target compound and 50 start orientations of the probe compounds is started by *COSMOsim3D*. The compound list consists of 88 compounds which are located in the subdirectory `examples/thr-ligands` of the installation directory.

To run this alignment job call `cs3d` on a shell (with ">" the screen output is redirected to a file):

- linux systems: `cs3d thr-align.inp > thr-align.log`
- windows systems: `cs3d.exe thr-align.inp > thr-align.log`

After `cs3d` has finished, the following files (in addition to the other example files) should be present alongside `thr-align.inp`:

`thr-align.out`: the output file

`thr-align.log`: the redirected screen output

`_tr.sms` files for all 88 compounds, i.e. `1_tr.sms`, `2_tr.sms`, `3_tr.sms`, ...

`_tr.cosmo` files for all 88 compounds, i.e. `1_tr.cosmo`, `2_tr.cosmo`, `3_tr.cosmo`, ...

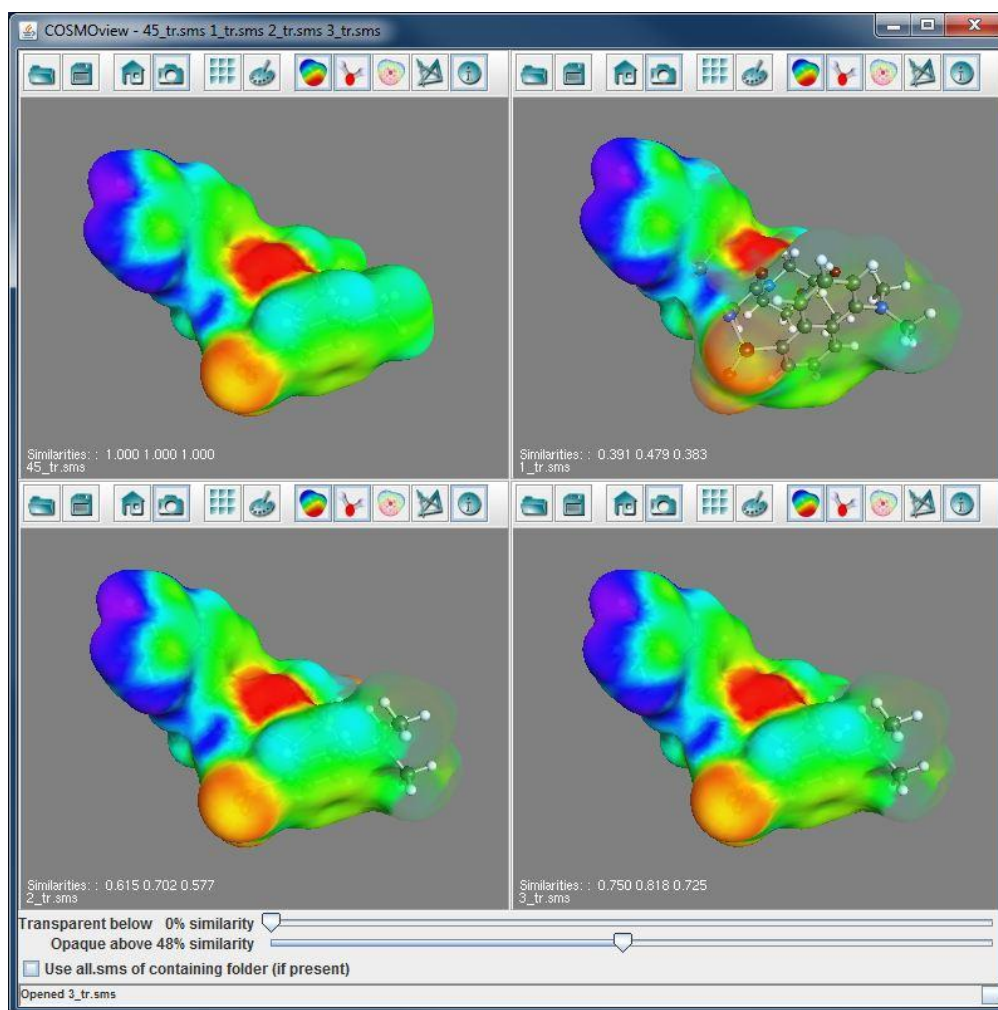
The `.sms` files can be visualized with *COSMOview* as described below.

## COSMOview visualization


The `.sms` files written by *COSMOsim3D* can be visualized with *COSMOview*, a free molecular structure viewer.

*COSMOview* can display one molecule per window or multiple frames with molecules in a multiframe mode. The multiframe mode is useful for visualizing aligned molecules in parallel. Rotation and zoom can also be done for all frames in parallel. The number of frames can be set with a right mouse button click on the box in the bottom right corner of the *COSMOview* window.


If `.sms` files are loaded, the window also features sliders to adjust the surface transparency based on the local `sms` values. The local `sms` values are used to control the COSMO surface transparency. With default settings, surface parts with high similarity are displayed with low transparency and intensive colors, while areas with lower similarity appear more transparent, so that the underlying molecular structure is visible.





The following options can be selected from the menu bar:

**OPEN** : Open one or more files from the file system. File types can be selected from the pull-down menu. To visualize a .sms file, select FILES OF TYPE: cosmo file (\*.wrl, \*.cosmo, \*.fcos, \*.ccf, \*.sms). If several files are selected and opened, COSMOview will automatically generate an appropriate multiframe window.


**SAVE** : Graphics can be saved, optionally with transparent background.


**RESET CAMERA** : Reset the camera to its initial position.


**LINK CAMERA** : In a multiframe window, apply camera movements (rotation and zoom) to all frames. By default, cameras in a multiframe window are linked.


**ARRANGE ALL WINDOWS** : The arrangement of multiple COSMOview windows can be changed.


**SETTINGS** : Change color, labels, atom settings, bond settings and so on.

**DISPLAY SIGMA SURFACE** : If a surface is loaded, it can be hidden and shown again.

DISPLAY MOLECULE : The molecule structure can be hidden and shown again.

DISPLAY WIREFRAME : Instead of closed object surfaces it is possible to show only the wire frame. This option works for surfaces, atoms and bonds.

USE CHARGE DENSITY PICKER : To get an idea of the quantitative surface charge density at a given point, you can activate the charge picking mode and move the cursor over the  $\sigma$ -surface. A slider at the right-hand side will display the charge density at the spot you are pointing on. However these values can only be approximated and are not guaranteed to be entirely precise. This is mainly an effect of interpolation between the reduced grid size compared to .cosmo files. Please also note that since COSMOview uses an internal color correction, the legend produced will not be applicable to images obtained by other means than COSMOview, e.g. third-party browser plug-ins.

DISPLAY INFO : Shows a few data on grid size and charge for

Movement: Molecules can be moved using the mouse buttons

Rotate the molecule by dragging the mouse with the left button pressed. If you move the mouse quickly, you can give the molecule a spin to have it turn by itself.

Zoom in and out with the right mouse button pressed or simply by turning the mouse wheel.

---

<sup>1</sup> A. Klamt, J. Reinisch, F. Eckert, A. Hellweg, M. Diedenhofen, Polarization charge densities provide a predictive quantification of hydrogen bond energies, *PCCP* **2011**, *14*, 955-963.

<sup>2</sup> M. Thormann, A. Klamt, M. Hornig, M. Almstetter, *COSMOsim*: Bioisosteric similarity based on COSMO-RS  $\sigma$ -profiles, *J. Chem. Inf. Model.* **2006**, *64*, 1040-1053.

<sup>3</sup> A. Klamt, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier: Amsterdam, **2005**.

<sup>4</sup> M. Hornig, A. Klamt, *COSMOfrag*: A novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. *J. Chem. Inf. Model.* **2005**, *45*, 1169-1177.

<sup>5</sup> P. Tosco, T. Balle, An open-source software aimed at high-throughput chemometric analysis of molecular interaction fields, *J. Mol. Model.* **2011**, *17*, 201-208.